Using Shannon Entropy to evaluate automatic music generation systems: A Case Study of Bach's Chorales

Iris Yuping Ren
ECE Department
University of Rochester
Rochester, NY 14623 USA
yuping.ren.iris@gmail.com

Abstract

In this position paper, we propose to use the Shannon Entropy (hereafter referred to as entropy) to perform self-evaluation in the automatic music generation systems. We demonstrate the process by analysing the entropy of pieces generated from two algorithmic composition systems ("EMI" (Cope 1992) and "Kulitta" (Quick 2014)) and the pieces from the actual style they try to emulate (J.S.Bach's chorales (Sapp 2005)). Using pitch-duration pairs, we found the two systems and the Bach's chorales have different levels of variety in their musical sequences. Eliminating the variety factor, the Kulitta system is different from the Bach's chorales in terms of repetition. We also found the pitch-duration pair entropy values of EMI's pieces are statistically higher than the composition of Bach, which indicates EMI is less predictable and this is very likely caused by the high variety level. Finally, we make the argument that this evaluation approach can be used widely with algorithmic stylist compositions to check whether the complexity of the generated music is in accordance with the original style.

Introduction

WIth the growing number of algorithmic composition systems, the evaluation of such systems is getting more attention (Papadopoulos and Wiggins 1999; Pearce and Wiggins 2001; Collins et al. 2016) yet not complete. Although it was reasonably argued that the most desirable way to evaluate algorithmically composed music is to have subjects listen to the music, it can be time-consuming and expensive to setup the experiments and not the best way to evaluate the system during the generation process. Instead, can we calculate some essential features from the original music and compare them with the computer generated music to have faster feedbacks?

The key is at whether we can find a credible feature to make the comparison. Entropy, a measurement of the unpredictability/complexity in a sequence of data, is a good candidate. The complexity of music has two aspects, the range of notes and the repetitiveness of the notes. Entropy values, on the other hand, are affected by the variety and the repetition in a sequence, too. Therefore, using entropy, we can capture the levels of variety and frequencial aspects of repeated note patterns in music, which is important musically. In fact, people have used entropy for analysing music in multiple occasions (Manzara, Witten, and James 1992; Dubnov and Assayag 2002; Laney, Samuels, and Capulet 2015), but as far as we know, there has not been research using the entropy measure to evaluate computer generated music, in particular, the "EMI" system (Cope 1992) and the "Kulitta" system (Quick 2014), which have achieved producing music credibly in the style of Bach's chorales according to (Johnson 1997; Devin 2015). In this paper, taking the compositions of these two systems, we further quantitatively investigate into how they are subtly different from the original Bach's chorales using the entropy.

Algorithmic Composition

The idea of algorithmic composition has been there for a long time: Mozart's "Musikalisches Wrfelspiel" (musical dice game) for example. In more modern times, famous composers such as Arnold Schoenberg, Anton Webern, John Cage all have composition based on different algorithms. However, it was not possible to massively and systematically produce these pieces until the recent rise of the computational power, especially for those algorithms which are imitating a specific genre since they usually require a training process on a large data set. There have been many systems designed to simulate the music creativity via mathematical models, knowledge-based systems, evolutionary methods, etc., and these methods have been applied to different genres and purposes of music creation: classical, jazzy, improvisational, modernism, atonal music, etc.. Despite such variety, to not be lack of comparability in this research, we chose a very specific style: J.S.Bach's chorales, since there are two famous and successful systems compose in this style, and the original corpus is available online as well.

For the data of the algorithmically composed music, we got EMI's music from (Cope 1989). The creator of this system, Prof. David Cope, placed 5000 MIDI files of computer-created Bach-style chorales for download, which was very convenient for our purposes. The Kulitta data was obtained by contacting the author directly. For the real Bach's chorales, we got data from (Sapp 2005). A hundred pieces from each system are used to make the comparison.

The Kulitta system generates music by using the probabilistic temporal graph grammars (PTGGs), while EMI is a based on a set of recombination instructions. For simplicity reasons and demonstrate the possibility of more generic use, we will not dive into more details of the algorithms but statistically analyse the entropy of the pieces created by the algorithms.

Information Theory

Mathematically, entropy is defined as in Equation 1 for a discrete probability space. It is the expectation value of the information content, which is defined as $-logp(x_i)$.

$$H(X) = -\sum_i p(x_i) logp(x_i), i \in n = outcomes \quad (1)$$

Intuitively speaking, we get more information from an unlikely event, and therefore an event with low probability gives us more information. For example, if a string is $11111...$, from the distribution $P(1) = 1$, we do not get much information from the next number, which is bound to be 1; the string $15342...$, from the distribution $P(x) = 1/n, x \in 1, 2, 3..., n$, on the other hand, is something we do not know much about its pattern and will not be able predict the next occurrences, and therefore gave us more information with respect to the knowledge we already know. Another interpretation from this is that, a high entropy indicates a surprisal element (get to maximum when the input follows uniform distribution), while a low entropy indicated a more predictable pattern (get to minimum when the input follows constant distribution).

In the case of music, we can calculate the entropy of a piece by counting the frequency of musical events. For example, we can count the appearances of each note(pitch-duration pair), pitch alone, duration alone and get the discrete distribution of those musical events in the piece. Then we can use the equation to calculate the information content of each note and then take the expectation to obtain the entropy. Essentially, as we discussed in the introduction, the entropy is tied with the frequency of musical events in a specific range. The differences in entropy values stem from the differences of, first, the underlying possibility space size, i.e. how many different types of musical events there are, second, how repetitive they are. Although this process is taking out of the consideration of the events' order, it characterise the complexity of music.

To introduce more properties on the relationship between entropy and repeating patterns, we introduce some properties as follows:

- For uniform distributions, the entropy increase with the number of outcomes, that is, the size of the underlying possibility space. For example, the entropy of a monotone note sequence is lower than the entropy of a scale.
- For discrete systems, the uniform distribution can be used as a baseline in comparing the entropy of different distributions, since it gives the maximum entropy value given a fixed possibility space size.
- The entropy is lower when there is reduced uncertainty. For example, tonal music has smaller entropy than atonal music, since there are more frequent note in the tonal, and this give us a smaller term in the definition of entropy.

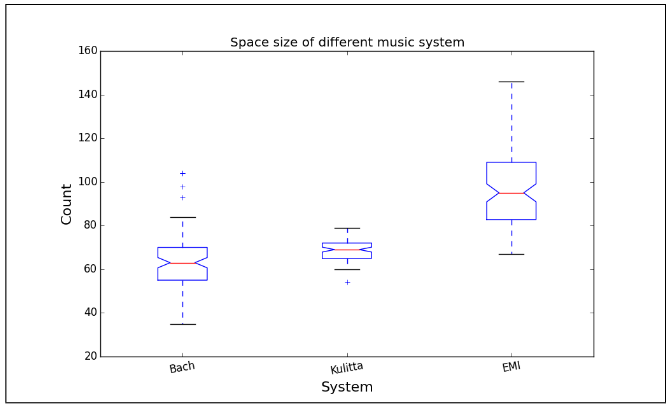

Figure 1: The underlying possibility space size of the pitch-duration pair of the three systems. Kulitta is narrower than the original and EMI is wider than the original. Each box plot is calculated using 100 pieces of music.

- The entropy remains the same when there is a repeat. For example, if we concatenate 2 bars of the same music together, the 4 bars music will have the same entropy as the 2 bars music.

The proofs of those properties are trivial from the definition of the entropy.

Results

To provide a common ground on the analysis of the repetitiveness, we first calculate the size of the possible pitch-duration space for each piece and summarise them using the box plot in Figure 1. We can see that, across the compositions of Kulitta, the pitch-duration combinations do not have as much variations as the original Bach and EMI. Also, we establish the fact that there is a difference in the range of notes different systems take. Therefore, we need a baseline to compare with to be conclusive about the repetition in music.

In Figure 2, the box plots show the entropy differences between the three systems and their baseline uniform distributions. Just by comparing the three systems, we can see that Kulitta has a similar range of entropy as Bach's chorales, yet EMI has a significantly higher entropy than the other two. This alone informs as that the complexity of EMI system's music is higher than the Kulitta system and Bach's chorales, but we cannot be conclusive about if this is due to a high level of variety or the lack of repetitiveness in music.

Looking at all six data points give us more information on the repetitiveness. As introduced in the last section, using the number of different musical events, we can obtain the maximum entropy by imposing an uniform distribution, which can serve as a baseline for reasoning on the repetitiveness of the music. We can see that the Kulitta system does not fit into the pattern formed by the other two: a big overlap in the actual entropy values and the maximum entropy values. This suggests that the Kulitta system is perhaps generating music with too much repetitive pitch-duration pairs.

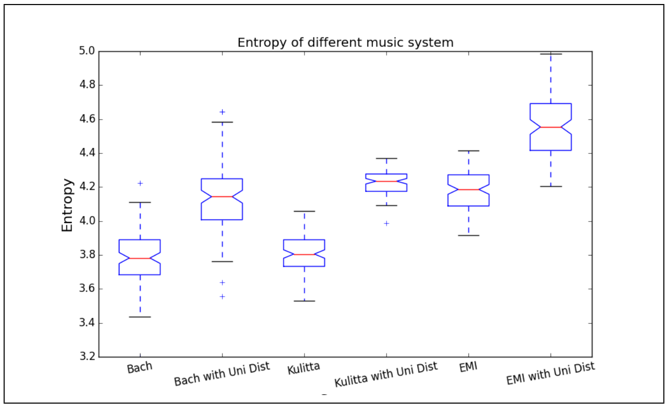

Figure 2: Entropy of pitch-duration pair compared with the baseline of the uniform distribution. EMI has a significanly higher entropy than the other two. The baseline/entropy relation of Kulitta system is different from the original and EMI. Each box plot is calculated using 100 pieces of music.

Summary

In summary, we first reasoned that calculating entropy of music has important meanings, and then gave backgrounds on algorithmic composition and information theory, followed by the calculation of entropy on the human-like algorithmically composed music (EMI's and Kulitta's chorales in Bach's style) and the actual human music (J.S.Bach's chorales). From the calculation and visualisation, we found instructive differences between the two, and then gave some suggestions as to how the systems might be able to improve.

We also demonstrated that this type of evaluation, using computational measures to compare the algorithmically composed music and the original, is easy to perform yet musically informative. Evaluating music generating algorithm this way could give a unique and quantitative sight as to how the system deviates from the original style and further improve the system. For example, when the entropy is too high, we can either reduce introducing new musical events or including more repeated patterns and prominent notes or both.

Future Works

For future works, first step would be to try implement the suggestions given by the analysis with the authors of the algorithms, and make listening tests to see if there are improvement. Furthermore, we could try this method with more data from more algorithmic composition systems and the corresponding comparable human music. There is also space for improving the entropy calculation: instead of just taking the notes probability distribution from each piece, to further mimicking the cognitive process of a real music listener, we could incorporate an underlying distribution of the listeners' expectation gained from their musical exposures. Moreover, one fundamental problem with the application of entropy is that we do not have the order of the notes considered, since entropy is defined only via distribution of occurrences, therefore ignoring which notes comes first and which comes after. To incorporate the order element, we can perhaps look at the entropy of repeated musical patterns instead of the individual notes. Finally, this line of thinking can go beyond the information theory: we could develop a more complete sets of methods to evaluate music-generating algorithms. Using these tools, only with very low cost, we could improve composition algorithms and make the gap between human-like algorithmically composed music and human music smaller in a unique and efficient way.

References

Collins, T.; Laney, R.; Willis, A.; and Garthwaite, P. H. 2016. Developing and evaluating computational models of musical style. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 30(01):16–43.

Cope, D. 1989. 5000 works in Bach style. http://artsites.ucsc.edu/faculty/cope/5000.html.

Cope, D. 1992. Computer modeling of musical intelligence in emi. *Computer Music Journal* 16(2):69–83.

Devin, C. 2015. Is it a computer or is it Bach? program fools even music experts. http://www.nbcnews.com/tech/innovation/computer-generates-music-rival-bachs-n414066.

Dubnov, S., and Assayag, G. 2002. Universal prediction applied to stylistic music generation. In *Mathematics and music*. Springer. 147–159.

Johnson, G. 1997. Undiscovered Bach? no, a computer wrote it. http://www.nytimes.com/1997/11/11/science/undiscovered-bach-no-a-computer-wrote-it.html?pagewanted=all.

Laney, R.; Samuels, R.; and Capulet, E. 2015. Cross entropy as a measure of musical contrast. In *Mathematics and Computation in Music*. Springer. 193–198.

MacKay, D. J. 2003. *Information theory, inference and learning algorithms*. Cambridge university press.

Manzara, L. C.; Witten, I. H.; and James, M. 1992. On the entropy of music: An experiment with Bach chorale melodies. *Leonardo Music Journal* 81–88.

Papadopoulos, G., and Wiggins, G. 1999. Ai methods for algorithmic composition: A survey, a critical view and future prospects. In *AISB Symposium on Musical Creativity*, 110–117. Edinburgh, UK.

Pearce, M., and Wiggins, G. 2001. Towards a framework for the evaluation of machine compositions. In *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, 22–32. Citeseer.

Quick, D. 2014. *Kulitta: A Framework for Automated Music Composition*. Ph.D. Dissertation, YALE UNIVERSITY.

Sapp, C. S. 2005. Online database of scores in the humdrum file format. In *ISMIR*, 664–665.